The Scope of Structural Isomerism[1]

DENNIS H. SMITH

Department of Chemistry, Stanford University, Stanford, California 94305

The variations in the number of structural isomers of organic compounds as a function of variation in atom type and degree of unsaturation[†] are discussed. These results provide quantitative measures which can be used to rationalize and extend intuitions of chemists on the scope of structural isomerism.

The fundamental concept of structural isomerism has been supported largely by intuitive ideas throughout the history of chemistry. Until recently there has been no systematic solution to the problem of specifying either complete sets of structural isomers for a given empirical formula or subsets based on constraints derived from a variety of sources (e.g., chemical isolation procedures, spectroscopic data). Problem-solving in this area, from student problems in introductory organic chemistry to complex problems of molecular structure elucidation, has been relegated to patient doodling with pencil and paper.

Central to the concept of structural isomerism is its scope, or the number *and* identity of each possible isomer.

This problem has not been ignored by chemists and mathematicians. A recent review of uses of graph theory in chemistry by Rouvray[2] summarizes past attempts to treat mathematically various problems of isomerism, with examples cited. With the exception of Lederberg's approach to generation of acyclic isomers[3] and recent work by Sasaki and coworkers[4] and Balaban,[5] the work summarized by Rouvray yields only the number, not the identities of isomers.[11]

The DENDRAL algorithm for generation of acyclic isomers[3] provides a systematic approach to the study of structural isomerism in acyclic molecules. More recently, an algorithm[7] and a computer program based on this algorithm[8] have been developed for exhaustive generation of structural isomers, inclusive of isomers containing cyclic and acyclic components. With the scope of structural isomerism amenable to treatment by systematic procedures, we can now consider detailed questions about isomerism.

[†] The term unsaturation, or degree of unsaturation, is used in this paper to mean the number of rings plus double bonds, or "double bond equivalents."
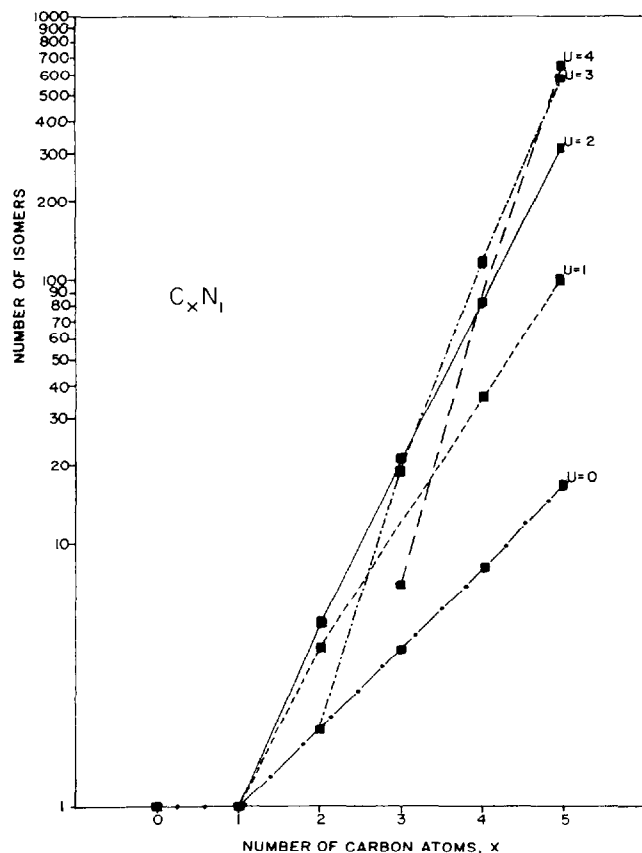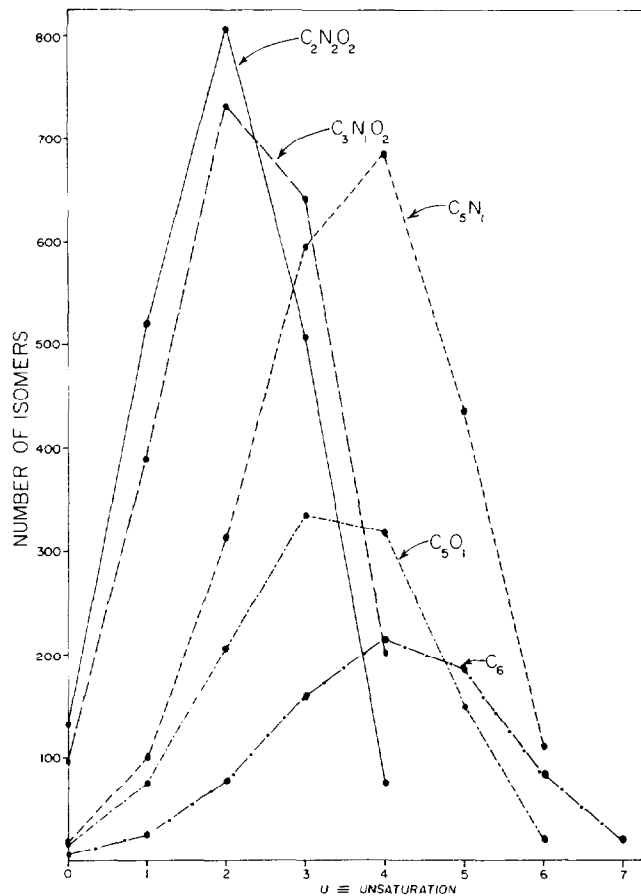
Figure 1. Semilog plots of the number of structural isomers vs. carbon number for empirical formulas containing 0–5 carbon atoms and one nitrogen atom and 0–4 U's (unsaturations, or double bond equivalents).

Figure 2. The number of isomers as a function of unsaturation count for groups of six atoms.

We selected a restricted set[12] of isomer generation problems, considering only a maximum of six nonhydrogen atoms, including carbon, nitrogen, and oxygen, to attempt to answer questions on the scope of structural isomerism which have arisen. These questions concern (1) the effect of increasing numbers of atoms on the total number of isomers; (2) the effect of increasing degree of unsaturation on the number of isomers for a given combination of C, N, and O; and (3) the effect on the number of isomers by replacing one atom type with another, retaining the same total number of atoms and unsaturations.

RESULTS

Each set of isomers was constructed by supplying to the structure generator[8] an empirical formula consisting of the number of each type of atom (carbon, nitrogen, oxygen) and the unsaturation count, which implicitly specifies the number of hydrogen atoms. The structure generator constructs the set of structural, or topological, isomers for the given formula. Thus, all isomers are constructed which obey normal valences ($C = 4$, $N = 3$, $O = 2$), but at present no account is taken of stereochemical considerations. Consideration of the three-dimensional properties of molecules would result in additional isomers. Neglect of bond lengths and angles and the inherent instability of certain functionalities (e.g., peroxides) results in construction of some structures which chemists would regard as implausible. However, what is implausible under some conditions may well be plausible under others. This paper does not attempt to define the chemical context within which certain structures may be forbidden; i.e., there is no constraint other than valence.

Use of normal values for valence means that certain functionalities which violate these values (e.g., nitro groups and N-oxides) are not considered in construction of structures. Consideration of them is straightforward, for example, by representing the nitro group as a univalent superatom,[8] $-NO_2$. The N-oxides are constructed using a trivalent superatom $O-N<$. This results in larger numbers of isomers than reported below for lists of atoms where such functionalities are possible. Finally, tautomeric structures are regarded by the program as distinct entities.

These caveats arise because we have a program which knows a great deal about graph theory, but very little about chemistry unless specific chemical constraints are supplied by the user (see Experimental Section). Within this restricted view of isomerism the results are quantitative. Additional considerations, such as stereochemistry, chemical context, and so forth, although they would modify the values presented below, do not significantly affect the conclusions.

The number of isomers for each combination of C, N, and O is presented in Table I as a function of the unsaturation count U. These numbers were determined by constructing each set of isomers and then counting the results.

The Number of Isomers vs. Increasing Number of Atoms. Chemists are aware that the scope of structural isomerism is immense. Yet it is our experience that educated guesses on the number of isomers for even small numbers of atoms frequently are an order of magnitude too low. Lederberg et al.[3a] have pointed out that the number of acyclic isomers increases approximately exponentially with increasing carbon number. As expected, semilog plots of the total number of structural isomers (data in Table I) display a similar behavior.[13] The data for $C_xN_1U_m$ ($0 \leq x \leq 5$, $0 \leq m \leq 4$, U = unsaturation) are presented in Figure 1. Most

such plots display a slightly decreasing slope with increasing carbon number for each given unsaturation.

As unsaturation increases, the slopes of the curves increase sharply (see Figure 1 as U increases from zero to four). However, curves representing higher unsaturation counts frequently begin, and remain for a short time, below curves representing lower unsaturation counts, for example, the curves for U = 3 and U = 4, Figure 1. It seems clear that the higher the degree of unsaturation the greater the possible number and combinations of multiple bonds and ring systems. This factor is responsible for increasing slope with increasing U observed in Figure 1. But, carrying this argument to its extreme, why should not the number of isomers for a given combination of atoms always be greater for higher unsaturation counts, as long as the total valence of the composition list can support the specified unsaturation? The next section discusses this point in more detail.

The Number of Structural Isomers vs. Unsaturation Count. In Figure 2 some selected data from Table I are plotted including various combinations of six atoms. This figure illustrates the typical rise and fall of the number of isomers with increasing unsaturation. The value of unsaturation at which the number of isomers reaches a maximum is a complex function of the number and identity of atoms. The value shifts slowly to higher values of unsaturation as the number of atoms increases. Most composition lists of four atoms show maxima at U = 1 or 2. Several composition lists of six atoms have maxima at U = 3. However, most still display maxima at U = 1 or 2.

It is interesting to note that there is a smooth variation of number of isomers with unsaturation. No example presented in Table I shows a number of isomers which declines, then rises again with increasing unsaturation.

The curves presented in Figure 2 serve to introduce another set of questions and observations. These are concerned with the effects on the number of isomers caused by replacement of one type of atom with another type.

The Number of Structural Isomers vs. Atom Type. Most chemists intuitively expect that a given number of atoms and unsaturations will yield a larger number of isomers with a more diverse collection of atoms. As a first approximation, this is true, as indicated in Figure 2. The curve of C_5N_1 lies above the curve of C_6 for all values of U; the curve of C_5O_1 lies above that of C_6 for $0 \leq U \leq 4$. It is reasonable to expect that, in a comparison of the results of exchange of atom types (e.g., either one N or one O for one C), the exchange yielding a higher total valence results in a group of atoms capable of yielding a greater diversity of structures. To a first approximation this is also true, as the curve of C_5N_1 lies above that of C_5O_1 for all values of U (Figure 2).

Results presented previously, however, indicate the subtle interplay of atom type and valence. Intuition often fails to predict the relative numbers of isomers resulting from replacing an arbitrary number of atoms by the same number of atoms which differ in name and valence. Some representative data selected from Table I are presented in Table II to indicate the influence of variation of atom type on the number of isomers.

For a given unsaturation count, the ratio between numbers of isomers resulting from N-substitution vs. O-substitution increases with increasing substitution, e.g., N_1/O_1, N_3/O_3, Table II.

DISCUSSION

The variations of numbers of isomers with increasing numbers of atoms, with unsaturation count or with substitution of one atom type with another can be rationalized on the same basis. Knowledge of an algorithm for constructing structural isomers, such as that used by the structure gen-

Table I. Structural Isomers of $C_xN_yO_z$, $x + y + z \leq 6$, for Allowed Values of the Unsaturation, U

No. of atoms	Unsaturation							
	0	1	2	3	4	5	6	7
C_1	1							
C_2	1	1	1	1				
C_3	1	2	3	2	1			
C_4	2	5	9	11	7	3		
C_5	3	10	26	40	40	21	6	
C_6	5	25	77	159	217	185	85	19
O_1	1							
C_1O_1	1	1						
C_2O_1	2	3	3	1				
C_3O_1	3	9	13	9	2			
C_4O_1	7	26	55	62	36	7		
C_5O_1	14	74	205	337	318	151	21	
O_2	1	1						
C_1O_2	2	2	1					
C_2O_2	5	10	9	3				
C_3O_2	11	34	52	34	7			
C_4O_2	28	122	263	301	163	28		
O_3	1	1						
C_1O_3	3	4	1					
C_2O_3	10	22	20	5				
C_3O_3	28	102	152	98	16			
O_4	1	1						
C_1O_4	5	6	2					
C_2O_4	20	48	41	10				
O_5	1	1						
C_1O_5	6	9	2					
O_6	1	1						
N_1	1							
C_1N_1	1	1	1					
C_2N_1	2	4	5	2				
C_3N_1	4	12	21	19	7			
C_4N_1	8	35	85	116	87	27		
C_5N_1	17	100	313	593	685	437	112	
N_2	1	1	1					
C_1N_2	2	4	4	1				
C_2N_2	6	18	27	19	5			
C_3N_2	14	62	136	155	86	14		
C_4N_2	38	218	633	1058	1005	465	64	
N_3	1	2	1					
C_1N_3	4	11	13	6				
C_2N_3	14	58	110	99	34			
C_3N_3	45	259	681	969	706	194		
N_4	2	4	4	2				
C_1N_4	8	29	47	31	6			
C_2N_4	37	189	439	512	272	42		
N_5	2	7	9	4				
C_1N_5	15	73	145	131	42			
N_6	4	15	25	18	6			
N_1O_1	1	1						
$C_1N_1O_1$	3	5	3					
$C_2N_1O_1$	8	22	26	11				
$C_3N_1O_1$	21	84	154	136	46			
$C_4N_1O_1$	56	299	764	1069	775	216		
N_2O_1	2	3	1					
$C_1N_2O_1$	8	21	18	4				
$C_2N_2O_1$	31	115	177	114	20			
$C_3N_2O_1$	102	527	1194	1371	703	88		
N_3O_1	3	7	4					
$C_1N_3O_1$	21	71	86	34				
$C_2N_3O_1$	101	481	935	826	256			
N_4O_1	6	18	17	4				
$C_1N_4O_1$	52	225	361	227	32			
N_5O_1	11	43	55	22				
N_1O_2	2	2						
$C_1N_1O_2$	8	15	8					
$C_2N_1O_2$	28	84	99	40				
$C_3N_1O_2$	90	391	732	641	202			
N_2O_2	5	8	3					
$C_1N_2O_2$	28	75	65	12				
$C_2N_2O_2$	132	521	807	506	76			
N_3O_2	10	23	13					
$C_1N_3O_2$	86	306	369	137				
N_4O_2	24	72	66	14				
N_1O_3	3	3						
$C_1N_1O_3$	17	34	18					
$C_2N_1O_3$	76	246	288	110				
N_2O_3	9	15	5					
$C_1N_2O_3$	73	207	173	29				
N_3O_3	24	58	30					
N_1O_4	4	4						
$C_1N_1O_4$	33	68	34					
N_2O_4	17	27	8					
N_1O_5	5	5						

Table II. Variation in the Number of
Isomers by Substitution of Nitrogen and/or
Oxygen Atoms for Carbon Atoms. Some Composition
Lists of Six Atoms with Two Unsaturations

Atoms, C_nXU_2		No. of
n	X	isomers
6		77
5	O_1	205
5	N_1	313
4	O_2	263
4	N_2	633
4	N_1O_1	764
3	O_3	152
3	N_3	681
3	N_1O_2	732
3	N_2O_1	1194

erator,[8] promotes a systematic way of thinking about isomerism which can be used in this reasoning.

The structure generator considers all possible ways of apportioning elements of the empirical formula, including atoms and unsaturations, among sets used to construct rings and corresponding sets used to construct acyclic chains which contain no unsaturation. The greater the number and variety of atoms, the greater the number of ways these sets may be assembled and the greater the potential number of isomers. Ever increasing numbers of unsaturations serve to increase the variety of ring systems which can be constructed, but at the expense of the variety of acyclic chains. Eventually atoms must be, in a sense, removed from the set of atoms in acyclic chains in order to support the additional number of unsaturations allocated to atoms in ring systems. But this is still only a partial explanation. Exactly what is serving to restrict the number of isomers at high degrees of unsaturation?

The answer lies in a broader conception of the term "variety" in reference to atoms. Curves of the number of isomers of n carbon atoms as a function of increasing unsaturation rise and fall even though the variety of atom names (e.g., all C's) remains constant. But if variety is extended to refer to atom names *and* the potential variety of degree of atoms of the same name, these curves make sense. At low values of unsaturation, most carbon atoms will be primary or secondary[14] and variety is low. As U increases, variety increases, as carbon atoms will exist in several different combinations of degree one through four in the various isomers. As U approaches the maximum which can be supported by the atoms, most carbon atoms will be quaternary, few or none primary or secondary, and variety has again diminished and the number of isomers has declined. This argument is expressed in terms of unsaturations and degree. One can also frame it in terms of C's, N's, and O's associated with hydrogen atoms rather than unsaturation. The argument then becomes the following, using carbon as an example: at low values of unsaturation, one is dealing primarily with methyl (CH_3) and methylene ($-CH_2-$) groups and variety is low; at intermediate values of unsaturation, there are more combinations because of the greater number of ways to distribute the hydrogen atoms among the various carbon atoms to produce several combinations of CH_3-, $-CH_2-$, $>CH-$, and $>C<$. At high unsaturation, there are few hydrogens to distribute among the carbon atoms, and most carbons are of the form $>C<$. (This argument is implicit in Balaban's discussion.[5])

Substitution of carbon atoms with other types of atoms also yields greater variety. Now, however, variety can be increased both by differing atom types and the potential for differing degrees of atoms of the same type. If substitutions are by atoms of the same valence, then the number of isomers will always be greater, for every value of U, than the number of isomers for the unsubstituted composition list. Thus, for any value of U, the number of isomers resulting from substitution of silicon for carbon will be in the order

(for example) $C_6 < C_5Si_1 < C_4Si_2 < C_3Si_3$ (but $C_2Si_4 = C_4Si_2$, variety is constant).

When the substitution is by an atom of lower valence, then variety of atom type is higher, but eventually the inability of the resulting group of atoms to maintain high variety at high degrees of unsaturation (due to lower total valence) may cause the number to fall below the results from a group of atoms of lower variety of atom type. These effects were mentioned previously (Figure 2). Note, in Figure 2, that C_5O_1 and C_5N_1 have the same variety of atoms, five C's, and one other. Yet the single extra valence of one nitrogen atom vs. one oxygen atom dramatically increases the variety based on consideration of the possible degrees of the atoms and, therefore, the number of isomers. Also, the composition lists $C_2N_2O_2$ and $C_3N_1O_2$ have greater variety of atom type than C_5N_1. Up to U = 2 (Figure 2) the former pair of composition lists yields more isomers than C_5N_1. At U = 4, however, total valence and its attendant restriction on the variety of degree of atoms (i.e., oxygen can only be of degree one or two, whereas nitrogen can possess one through three, and so forth) cause the number of isomers of $C_2N_2O_2$ and $C_3N_1O_2$ to fall below that of C_5N_1.

Similar arguments can be invoked to explain other features of the data presented in Table I. For example, numbers of isomers of $C_xN_2O_1$ and $C_xN_1O_2$ are consistently greater than those of C_xN_3, until high values of unsaturation are considered. The higher valence of C_xN_3 eventually supports a greater variety of degrees of atoms, and, thus, isomers. As expected, $C_xN_1O_2$ falls below C_xN_3 faster than $C_xN_2O_1$ as unsaturation increases.

CONCLUSIONS

Variations in the numbers of structural isomers as a function of several parameters have been presented and discussed. The concept of variety of atoms, whether with respect to degrees of atoms of the same name or with respect to differing atom names, is useful in thinking about these variations. No attempt to derive a quantitative measure of variety has been made. A quantitative measure would yield the number of isomers for a given composition list. There is a computational method[9] which yields this number, but this method is extremely time-consuming and does not, of course, determine the identity of each isomer.

A qualitative measure of variety is the number of initial partitions (superatom partitions[8]) of atoms and unsaturations among rings and chains. This, in fact, is the basis for the qualitative arguments. There is no easy formula known to us which can be used to calculate the number of superatom partitions. One must do the partitioning and count the results.

It is hoped that this discussion illustrates that there are procedures for systematic treatment of structural isomerism, and knowledge of these procedures can stimulate new ways of thinking about isomerism. Although the findings and rationalizations discussed are not surprising in hindsight, it would be difficult to predict or rationalize the aspects of the scope of structural isomerism discussed herein.

The isomer counts presented in Table I are useful primarily for the discussions which followed. The scope of isomerism is such that complete compilations, including counts and structures, of structural isomers are seldom useful. Chemists attack structure problems with many constraints on what constitute plausible structures. The discussion presented, however, can also assist in analysis of such constrained problems. A program (called "CONGEN") is available which is capable of exploring the scope of structural isomerism under constraints[10] (see Experimental).

EXPERIMENTAL

The structure generator CONGEN is written in the pro-

gramming language INTERLISP. Although we will provide copies and listings of the program to interested persons, this information will only be useful to a facility which maintains the INTERLISP language. The Stanford University Medical Experimental (SUMEX) computer facility has been established to encourage sharing of such programs among a collaborative community of users via a nationwide computer network. For additional information, write to the author, or to Professor Joshua Lederberg, Principal Investigator, SUMEX Project, Stanford University Medical School, Stanford, California 94305. Structures of isomers, including those presented here, with or without constraints, are available via this mechanism.

ACKNOWLEDGMENTS

REFERENCES AND NOTES

(1) Applications of Artificial Intelligence for Chemical Inference. XVIII. For part XVII, see ref. 10.

(2) Rouvray, D. H., *Chem. Brit.*, **10**, 11 (1974).

(3) (a) Lederberg, J , Sutherland, G. L., Buchanan, B. G., Feigenbaum, E. A., Robertson, A. V., Duffield, A. M., and Djerassi, C., *J. Am. Chem. Soc.*, **91**, 2973 (1969); (b) Buchanan, B. G., Duffield, A. M., and Robertson, A. V., "Mass Spectrometry: Techniques and Applications", G. W. A. Milne, Ed., Wiley, New York, N. Y., 1971, p 121.

(4) Sasaki, S., Kudo, Y., Ochiai, S., and Abe, H., *Mikrochim. Acta,* 726 (1971).

(5) Balaban, A. T., *Rev. Roum. Chim.,* **18**, 635 (1973).

(6) Polya, G., *Acta Math.,* **68**, 145 (1937).

(7) Brown, H., and Masinter, L. M., *Discrete Math.,* **8**, 227 (1974).

(8) Masinter, L. M., Sridharan, N. S., Lederberg, J., and Smith, D. H., *J. Am. Chem. Soc.,* **96**, 7702 (1974).

(9) Masinter, L. M., unpublished results.

(10) Carhart, R. E., Smith, D. H., Brown, H., and Djerassi, C., *J. Am. Chem. Soc.,* in press.

(11) Rouvray[2] uses the term "enumeration" in reference to both the constructive approach of Lederberg[3] and the isomer counting technique based on Polya's theorem.[6] The former technique yields the identity of each isomer, information we deem critical to chemists; the latter technique does not. Thus, we refer to structure generation, or construction, with reference to methods which yield the actual structures.

(12) This set of problems is restricted for several reasons. Selected isomer counts are presented as they are important to the discussion. Practicality dictates against other, larger problems; they would represent only extensions to the discussion. The program is not restricted to the problems presented. Structures from these problems and others are available (see Conclusions and Experimental).

(13) Disagreements between the first column (U = 0) of Table I and Table III of reference 3a arise from the fact that some undesired substructures (BADLIST[3]) were not considered in the latter tabulation.

(14) We use the terms primary, secondary, . . . , in reference to a broad definition of the word "degree". We consider for the purposes of this discussion that a tertiary carbon atom possesses three bonds to other, nonhydrogen atoms, independent of the multiplicity of these bonds. Although exception might be taken to this broad a definition, it is consistent with the structure generator's topological view of chemistry and saves having to introduce the textual complexities of differing hybridizations. We only note that the different hybridizations which, for example, a quaternary (by our definition) carbon may possess lead also to greater variety, which is responsible for there being more isomers of C_6U_7 (no hydrogens) than C_6U_0. Although the latter composition list can support carbon atoms with several degree values (e.g., 2,2-dimethylbutane has carbon atoms of degree 1, 2, and 4), they are all sp^3 carbons. The former composition list has only quaternary carbon atoms, but there are several different hybridizations possible.